The Cartesian argument against physicalism

Philip Goff

I'm an analytic metaphysician who thinks analytic metaphysicians don't think enough about consciousness. By consciousness I mean the property of being a thing such that there's something that it's like to be that thing. There's something that it's like for a rabbit to be cold, or to be kicked, or to have a knife stuck in it. There's nothing that it's like (or so we ordinarily suppose) for a table to be cold, or to be kicked, or to have a knife stuck in it. There's nothing that it's like from the inside, as it were, to be a table. We mark this difference by saying that the rabbit, but not that table, is conscious.

The property of consciousness is special because we know for certain that it is instantiated. Not only that but we know for certain that consciousness *as we ordinarily conceive of it* is instantiated. I am not claiming that we know everything there is to know about consciousness, or that we never make mistakes about our own conscious experience. My claim is simply that one is justified in being certain—believing with a credence of 1—that there is something that it's like to be oneself, according to one's normal understanding of what it would be for there to be something that it's like to be oneself.

This makes our relationship with consciousness radically different from our relationship with any other feature of reality. Much metaphysics begins from certain 'Moorean truths'; truths of common sense that it would be intolerable to deny. Perhaps it is a Moorean truth that some or all of the following things exist: persons, time, space, freedom, value, solid matter. But it would be difficult to justify starting metaphysical enquiry from the conviction that these things must exist as we ordinarily conceive of them. We must remain open to science and philosophy overturning our folk notions of what it is for someone to be free, or for something to be solid, or for time to pass.

Matters are different when it comes to consciousness. It is not simply that I can gesture at some property of 'consciousness' with folk platitudes, and have confidence that something satisfies the bulk of those platitudes. When I entertain the proposition <there is something that it's like to be me>, I know that *that very proposition* (not it or some revision of it containing a slightly different concept of 'being something such that there's something that it's like to be it') is true.

You can't build a satisfactory metaphysical theory wholly from the datum that there is consciousness; that datum is after all consistent with solipsism. We must continue to rely on Moorean truths, empirical data and the weighing of theoretical virtues in trying to formulate our best guess as to what reality is like. But because the datum that there is consciousness (as we ordinarily conceive it) is unrevisable, it ought to occupy a central place in enquiry; a fixed point around which other considerations revolve. I call an approach to analytic metaphysics that grants the reality of consciousness this central place 'analytic phenomenology'.

The potential of this datum is grossly underexplored; it has arguable implications for the nature of time, persistence, properties, composition, objecthood and personal identity. Time will tell, but it is possible that, with an agreed source of unrevisable data, analytic phenomenologists may achieve some degree of consensus on certain key questions, a goal which has so far eluded other schools of metaphysics.

Perhaps the most famous alleged implication of the reality of consciousness is the falsity of physicalism. In what follows I will present an argument against standard physicalist attempts to account for consciousness. In my undergraduate lectures, Descartes' arguments against materialism were presented as objects for target practice rather than serious evaluation. At the time it seemed to me that there was more to the arguments than they were being given credit for. I now think Descartes' *Meditations* provides us with the resources for a

sound argument against standard contemporary forms of physicalism. In what follows I shall present this argument.

In the final section, I will highlight a distinctive advantage of this argument: if sound, it demonstrates the non-physicality not only of sensory experience, but also of thought.

1. The second meditation and the refutation of analytic functionalism

Physicalism is the metaphysical view that nothing in actual concrete reality is anything over and above the physical. There is a great divide amongst physicalists over the epistemological implications of that metaphysical doctrine. A priori physicalists, whom I shall consider in this first section, believe that all facts are a priori entailed by the physical facts. If you knew the intrinsic and extrinsic properties of every fundamental particle and field, and you were clever enough, you could in principle work out *a priori* all the other facts: what the chemical composition of water is, who won the Second World War, how many number one hits the Beatles had, etc.[1]

Perhaps the trickiest case for the a priori physicalist is mentality. Prima facie it doesn't seem possible to move a priori from the kind of facts brain science delivers to the facts about consciousness. A colour blind brain scientist could know all the physical facts about colour experience without knowing what it's like to see colours. A brain scientist who's never tasted a lemon could not work out how they taste from poking around in someone's brain.[2] Or at least that's how things seem. If she wants to have a plausible, fully worked out view, the a priori physicalist cannot just brutally assert that, contrary to appearances, the mental facts do follow a priori from the physical facts, but must give some plausible account of mental concepts which has this implication.

The standard way of doing this is to adopt some form of analytic functionalism, i.e. to give some kind of causal analysis of mental concepts. The straightforward analytic

functionalist says that mental concepts denote higher-order functional states, for example, the concept of pain denotes the state of having some more fundamental state that 'plays the pain role', i.e. (roughly) that responds to bodily damage by instigating avoidance behaviour. On the more subtle 'Australian' form of analytic functionalism, defended by David Armstrong and David Lewis, mental concepts are non-rigid designators which pick out certain states in virtue of the higher-level functional states they realise.[3] Just as the concept 'head of state' picks out in each country the individual that happens to be the head of state in that country, so 'pain' picks out in each population the state that happens to play the pain role in that population.

It is clear that both forms of analytic functionalism are forms of a priori physicalism. Suppose Jennifer's c-fibres are firing, and the firing of c-fibres is the state that plays the pain role both in Jennifer and in the human population in general. For the straightforward analytic functionalist, if I know all the physical facts I will be able to work out that Jennifer is in a state that plays the pain role, and can infer from this information that Jennifer is in pain. For the Australian analytic functionalist, if I know all the physical facts I will know that Jennifer instantiates the state that plays the pain role in humans, and can infer from this information that Jennifer is in pain. In either case, the mental facts can be deduced from the physical facts. The second meditation provides the resources for a decisive refutation of both of these forms of analytic functionalism. By the end of the second meditation I have doubted the existence of my body and my brain, and of the entire physical world around me. For all I know for certain, my apparent experience of all these things might be an especially vivid hallucination instigated by an omnipotent evil demon. This demon might have bought me into existence just a moment ago—with false memories of a long history and expectations of a similar future—and may destroy me a moment hence. I discover that the only thing the demon

cannot be deceiving me about is my own existence as a thinking thing: no matter how much the demon is deceiving me I must exist as a thinking thing in order to be deceived.

At the end of this guided meditation, when I have doubted the existence of anything physical whilst at the same time enjoying the certain knowledge that I exist as a thinking thing, I find I am conceiving of myself as *a pure and lonely thinker*: a thing which has existence only in the present moment, and that has no characteristics other than its present mode of thought and experience.[4] The fact that I can conceive of myself as a pure and lonely thinker is inconsistent with the analytic functionalist analysis of mental concepts. For the straightforward analytic functionalist, it is a priori that something has a given mental state if and only if it has the higher-order state of having some other state that plays the relevant causal role. However, a pure and lonely thinker has no states other than the mental states themselves: its mental states are not realised in anything more fundamental. If straightforward analytic functionalism is true, a pure and lonely thinker is inconceivable. And yet a pure and lonely thinker is not inconceivable; the second meditation guides us to its conception.

For the Australian analytic functionalist, it is a priori that something is in pain if and only if it has the state that plays the pain role in its population. But a pure and lonely thinker does not have a population; it is alone in its world. If Australian analytic functionalism were true, a pure and lonely thinker would be inconceivable. And yet by the end of the second meditation we end up conceiving of one.

Lewis does suggest at one point that the population relevant to determining the application of mental concepts might be the concept user's population, rather than the population of the creature the concept is being applied to.[5] However, when I reach the end of the second meditation, I am supposing that *I* am alone in the universe, and hence am not a member of any population. If Lewis were right about the reference fixing description of pain,

then ‘pain’ would have no application in such a conceivable scenario, just as concept ‘the head of state’ has no application in a scenario where there are no countries. And yet, if I read the second meditation when I have a headache, I end up conceiving of a scenario in which the concept ‘pain’ evidently has application.

Why have analytic functionalists been so complacent about this incredibly powerful argument against their view, an argument which—on the assumption that they took a philosophy degree—they cannot possibly have been ignorant of? I think that straightforward analytic functionalists have felt unthreatened by Cartesian considerations because their view entails that the mental is *multiply realised*, and allows that in certain non-actual scenarios the mental may be realised by non-physical goings on. Whilst functional states in the actual scenario may be realised by fleshy mechanisms, in non-actual scenarios they are realised by ectoplasm. Therefore, the fact that we can conceive of mental processes without physical processes—as we do at the end of the second meditation—is consistent with straightforward analytic functionalism.

However, whilst it is true that straightforward analytic functionalism is consistent with the conceivability of minds without brains, it is not true that straightforward analytic functionalism is consistent with the scenario we finish up conceiving of at the end of the second meditation. The thing we end of conceiving of at the end of the second meditation is not just a thing with mentality not realised in physical stuff, it is a thing with mentality not realised in *any* stuff. And it is not coherent to suppose that ‘the higher-order state of having some other state that plays the pain role’ exists in the absence of some other state that plays the pain role.

The Australian analytic functionalist avoids this problem by identifying pain with the realiser of the pain role rather than with the pain role itself. It is then conceivable that pain is a fundamental state, as there are scenarios where a fundamental state plays the pain role in

the population being considered. Furthermore, there are coherent scenarios where pain does not play the pain role. In cases of what Lewis calls 'mad pain' there is an individual who instantiates the state which plays the pain role in her population, without it being the case that it plays the pain role in her.[6] However, despite the ingenious flexibility of the Australian view, it does not allow that 'pain' has application in scenarios in which nothing in existence plays the pain role. And yet when we reach the end of the second mediation, and we have a headache, we find ourselves conceiving of a scenario in which nothing plays the pain role and yet 'pain' still has application.

In none of this discussion have we moved from the epistemological to the metaphysical. Analytic functionalists make certain claims about mental concepts, which have implications for what it is coherent to suppose. Those claims are inconsistent with the state of conceiving we end up in at the end of the second meditation. We are able to refute analytic functionalism by refuting its epistemological elements. Descartes admits in the second meditation that physical things—'these very things which I am supposing to be nothing, because they are unknown to me'—may 'in be reality identical with the 'I' of which I am aware'.[7] The leap from the epistemological to the metaphysical must wait until the sixth meditation.

2. The sixth meditation and the refutation of a posteriori physicalism

In the sixth meditation, we find the following argument against materialism:

> First, I know that everything which I clearly and distinctly understand is capable of being created by God so as to correspond exactly with my understanding of it. Hence the fact that I can clearly and distinctly understand one thing apart from another is enough to make me certain that the two things are distinct, since they are capable of

> being separated, at least by God. The question of what kind of power is required to bring about such a separation does not affect the judgement that the two things are distinct. . . . on the one hand I have a clear and distinct idea of myself, in so far as I am simply a thinking, non-extended thing; and on the other hand I have a distinct idea of body, in so far as this is simply an extended, non-thinking thing. And accordingly it is certain that I am really distinct from my body and can exist without it.[8]

We can lay the argument out as follows:

> **Premise 1**: Anything I can clearly and distinctly conceive of is possible.
>
> **Premise 2**: I can clearly and distinctly conceive of my mind and brain/body existing independently of each other
>
> **Conclusion 1**: My mind and my brain/body could exist independently of each other.
>
> **Premise 3**: If my mind and brain/body could exist independently of each other then they are distinct substances.
>
> **Conclusion 2**: My mind and brain/body are distinct substances.

Let us consider the premises of this argument in more detail.

2.1 Premise 1

When I was a first year philosophy undergraduate, I was taught that premise one of this argument could be swiftly refuted with the counterexample of water existing independently of H_2O. It seems that we can conceive of a scenario in which water exists in the absence of H_2O, for example a scenario in which experiments reveal water to have some other chemical composition. And yet if we infer from this the real possibility of water existing in the absence

of H_2O, we are quickly led to the non-identity of water and H_2O, contrary to what is in fact the case.

This rejection of premise one is far too quick. Descartes doesn't say that any old conceiving implies possibility, only that *a clear and distinct conception* implies possibility. I take it that whatever else having a clear and distinct conception involves, it involves *understanding what you're conceiving of*. Suppose I think of electric charge as 'that thing Dave (my physicist chum) was talking about the other night' (where I use this description as a rigid designator), but have zero understanding of the defining characteristics of negative charge. It is clear that such a conception of negative charge is not clear and distinct. I can refer to negative charge, but there is a clear sense in which I don't know what it is: I have no idea what it is for something to be negatively charged. I don't have the understanding of the nature of negative charge that—let us suppose—my physicist chum Dave has. Although I can involve negative charge in what I am conceiving of, to the extent that I do I bring opacity into my conception.

Such opacity brings in its wake coherent conceivability without possibility. I can coherently conceive of all sorts of scenarios in which 'negative charge' features—I might suppose that negative charge is what underlies a wizard's ability to teleport—without this implying that negative charge really could be as I am supposing. My ignorance of the nature of negative charge licences a conceptual free for all.

Our concept 'water' is also opaque in this sense. For something to be water is for it to be H_2O. But this is not apparent, or a priori accessible, to me when I conceive of water as such. Again we have a licence for a conceptual free for all: the fact that I can coherently conceive of water's having a chemical composition other than H_2O has no modal implications.[9] None of this takes us away from Descartes' view as to the relationship between conceivability and possibility. Perhaps Descartes had a false view about the nature of our ordinary concept of

water, but had he taken it to be an *opaque concept*, that is to say, a concept that reveals little or nothing about the nature of its referent, he would no doubt have denied that our conception of water when deploying that concept is clear and distinct, and hence denied that such conceptions have modal ramifications. On the assumption that a clear and distinct conception must involve only *transparent concepts*, that is, concepts that reveal the complete essence of the states they denote, the conceivably impossible scenario of water existing in the absence of H_2O does not constitute a counterexample to premise 1.

Indeed, once premise 1 is clarified in this way, it is not clear that there are any counterexamples to it. Putting aside the mind-body case as contentious, the examples one finds in the literature—individuals with origins distinct from their actual origins, or natural kinds with essences distinct from their actual essences—all seem to involve things being thought about under opaque concepts; in each case if we knew the essence or essential origins of the thing being thought about, the scenario in question would not be conceivable. Furthermore, there are clear benefits to the view that conceivability and possibility are linked in something like the way Descartes took them to be. It provides a clear and plausible account of how we know about possibility, and it offers the hope of an attractive reduction of modal truths in terms of facts about ideal conceivability (under transparent concepts).[10] Where it not for the trouble it makes for physicalism, perhaps this traditional view of the relationship between conceivability and possibility might not have fallen from favour.

2.2 Premise 2

When I reach the end of the second meditation, when I have stripped away everything it is possible to doubt and alighted upon the certain knowledge of my existence as a thinking/experiencing thing, I end up conceiving of my mind existing in the absence of anything physical. But is this conception clear and distinct? If either the concept of my mind,

or the general concept of the physical, is not fully transparent, then the resulting conception will fail to be clear and distinct.

Arnauld complained that Descartes had not demonstrated that our concepts of body and mind were *adequate*. Certainly they seem to reveal something of the nature of the substance they denote, but how can we know that they reveal its entire nature? Arnauld supports his argument by means of an analogy. Two things are worth noting about Arnauld's analogy: (i) it involves properties rather than substances (as Descartes notes in his reply[11]), (ii) it involves subtle a priori knowledge concerning those properties:

> Suppose someone knows for certain that the angle in a semi-circle is a right angle, and hence that the triangle formed by this angle and the diameter of the circle is right-angled. In spite of this, he may doubt, or not yet have grasped for certain, that the square on the hypotenuse is equal to the squares on the other two sides; indeed he may even deny this if he is misled by some fallacy.[12]

For this to be analogous to the mind-body case, the following would have to be the case. The reason we are able, at the end of the second meditation, to doubt the instantiation of physical properties without doubting the instantiation of mental properties, is that we have not worked out the subtle conceptual connection between the mental and the physical. On further reflection, it would turn out to be incoherent to suppose that, say, my current feeling of pain exists whilst nothing physical does, just as it turns out to be incoherent to suppose that the angle in a semi-circle is a right angle and yet the square of the hypotenuse (of a triangle formed from this angle) is not equal to the squares on the other two sides.

However, consider what such a subtle conceptual connection would involve. Nobody takes seriously the idea that there is a conceptual connection between mental properties and

specific physical properties, such as the firing of c-fibres. If this were the case, then neuroscience would be an a priori science. And so inevitably any supposed conceptual connection must be between mental properties and certain functional role properties, such that in the actual world those functional properties are realised by physical properties. If Arnauld's objection is to have any force, we must turn to analytic functionalism.

As we have seen, at the end of the second meditation we are conceiving of mental properties in the absence of such functional role properties, as we are doubting the latter but certain of the former. The analytic functionalist development of Arnauld's point would go as follows:

> We are only able to simultaneously suppose the existence of mental properties and doubt the existence of causal role properties because we have not reflected enough. Further reflection would reveal such a scenario to be incoherent.

However, it is not plausible to suppose that there is a conceptual connection between mental and causal role properties which it is too subtle to be noticed without some incredibly sophisticated reflection. We are not dealing with complicated mathematics here. Rather the analytic functionalist proposal is that causal role properties constitute the basic a priori content of mental concepts, that to suppose that someone is in pain just is to suppose that someone has an inner state that plays the pain role. If this were true, then at the end of the second meditation Descartes would be contradicting himself in the most perverse and straightforward way. This is simply not plausible.

Therefore, understood as a point about subtle conceptual connections between mental properties and physical or functional properties, Arnauld's concern has little force. However, there is still a serious issue concerning how Descartes can rule out that mental and physical

concepts fail to reveal the complete essences of the entities they denote. It is especially difficult to see how Descartes can rule this out concerning our concepts of mental and physical *substances*, as opposed to properties, which is of course the very kind of concepts he uses in his argument.

At the end of the second meditation, I am thinking of myself in terms of my mental properties. But how can I rule out that there is more to my nature than those mental properties I am using to think about myself? For the reasons I give above, it is implausible to suppose that there are some subtle corporeal aspects to my nature that are *conceptually implied* by the way I am conceiving of myself at the end of the second meditation. Nonetheless, there may be corporeal aspects of the 'I' I am conceiving of at the end of the second mediation which have no conceptual association with the mental properties in terms of which I am conceiving of that 'I'. For this reason, it seems to me that Descartes' argument fails as an argument for substance dualism, on the grounds that Descartes is unable to demonstrate that the concept of myself I have at the end of the second meditation reveals my complete nature, and hence is unable to demonstrate that premise two is true.

However, I want to suggest that the argument can be modified to form a successful argument for property dualism, by substituting the following for premise 2:

> **Premise 2***: I can clearly and distinctly conceive of my mental properties existing in the absence of any neurophysiological or functional properties.

and the following for premise 3:

> **Premise 3***: If my mental properties could exist independently of any neurophysiological or functional properties, then my mental properties are not identical to any neurophysiological or functional properties.

When I reach the end of the second meditation, I am conceiving of my mental properties existing in the absence of any physical or functional properties. But is this a clear and distinct conception? Could it be that the properties denoted by my mental concepts are in fact physical or functional properties, even though this is not apparent a priori? Or could it be that the properties denoted by my physical properties are in fact mental properties, even though this is not apparent a priori? Let us take both of these possibilities in turn.

2.3 Mental concepts

These days analytic functionalism isn't so popular, and most philosophers of mind accept the existence of distinctively mental concepts which bear no a priori connection to physical or functional concepts. However, physicalists tend to embrace a semantic externalist account of the reference of such concepts. The reference of our mental concepts, on such a view, is determined by facts outside of what is a priori accessible: causal connections, sub-personal recognitional capacities of the concept user, or facts about the evolved function of our mental concepts.[13] If the reference of our mental concepts is determined by facts outside of what is a priori accessible, then mental concepts lack a priori content, they are utterly opaque 'blind pointers'. Just as we pick out water as 'that stuff whatever it is' (pointing with our fingers), so we pick out pain as 'that state, whatever it is' (pointing introspectively). It turns out, thinks the physicalist, that in each case 'that state whatever it is' is a brain state.

If this kind of view is correct, then we do not have a clear and distinct conception of our mental properties when we think of them under mental concepts, and premise 2* is false. But such a view of our mental concepts is utterly implausible. Perhaps the best way to see the implausibility is to return to the second meditation. Recall how one ends up conceiving of oneself at the end of the second meditation:

> . . . what then am I? A thing that thinks. What is that? A thing that doubts, understands, affirms denies, is willing, is unwilling, and also imagines and has sensory perceptions [by 'sensory perceptions' Descartes conscious experiences as though perceiving through the senses].[14]

On the semantic externalist model, each of the mental concepts involved in this description is a blind pointer, revealing nothing about what it is for something to be in the state denoted. But it is evident when I am in it that the conception I have of myself at the end of the second meditation, when I have doubted away the physical world, is a rich, substantive conception of myself. If I know of someone (myself, or someone else) that they are doubting such and such, or understanding such and such, or believing or wanting such and such, or that they are having certain sensory experiences as though such and such were the case, I understand a great deal about that person's nature. I am not just blindly denoting their qualities.

Gassendi seems to worry in one of his objections to Descartes that to conceive of himself as a mental thing is not to have a substantive conception of his nature:

> Who doubts that you are thinking? What we are unclear about, what we are looking for, is that inner substance of yours whose property is to think. Your conclusion should be related to this inquiry, and should tell us not that you are a thinking thing, but what sort of thing this 'you' who thinks really is. If we are asking about wine, and looking for the kind of knowledge which is superior to common knowledge, it will hardly be enough for you to say 'wine is a liquid thing, which is compressed from grapes, white or red, sweet, intoxicating' and so on. You will have to attempt to investigate and somehow

> explain its internal substance, showing how it can be seen to be manufactured from spirits, tartar, the distillate, and other ingredients mixed together in such and such quantities and proportions. Similarly, given that you are looking for knowledge of yourself which is superior to common knowledge (that is, the kind of knowledge we have had up till now), you must see that it is certainly not enough for you to announce that you are a thing that thinks and doubts and understands, etc. You should carefully scrutinise yourself and conduct, as it were, a kind of chemical investigation of yourself, if you are to succeed in uncovering and explaining to us your internal substance.[15]

Descartes replies 'I have never thought that anything more is required to reveal a substance than its various attributes; thus the more attributes of a given substance we know, the more perfectly we understand its nature.'[16] I would want to make a slightly more qualified claim: all that is required to reveal a substance is knowledge of its attributes *under transparent concepts*. The description Gassendi offers of wine is formed of opaque concepts, which fail to tell us the real nature of wine: 'wine is a liquid thing, which is compressed from grapes, white or red, sweet, intoxicating.' In such a case, empirical investigation is required to make progress on understanding the nature of wine.[17] Gassendi has correctly identified some common or garden concepts which happen to be opaque. But it of course does not follow that *all* of our common or garden concepts are opaque. Whether or not mental concepts are fully transparent I shall consider shortly. But it is evident that the conception of myself I have at the end of the second meditation is not entirely opaque; it reveals to me at least something of my nature.

The second meditation alone, then, provides the resources to refute both the dominant form of a priori physicalism and the dominant form of a posteriori physicalism. Once

Descartes has guided me to a conception of myself as a pure and lonely thinker, it is evident that:

A. I can coherently suppose that nothing exists other than myself and my conscious experience, from which I can infer that analytic functionalism is false.
B. I am having a rich and substantive conception of my nature, from which I can infer that the semantic externalist model of mental concepts favoured by most contemporary physicalists is false.

It is worth noting that both of these positions can be ruled out without moving from the epistemic to the metaphysical. These specific physicalist views make specific claims about mental concepts which are inconsistent with the fact that the conception I have of myself at the end of the second meditation is substantive and consistent. However, if we want to refute a more general conception of physicalism, rather than specific (albeit very widespread) versions of it, we must try to make that move from conceivability to possibility by justifying premise 2* or something like it.

Our mental concepts are certainly not opaque, but does it follow that they are transparent, revealing everything of the nature of the mental properties they denote? Perhaps rather they are *translucent*, revealing some but not all of the nature of mental properties. Although in conversation many philosophers seem attracted to the idea that mental concepts are translucent, there are not many worked out versions of the view. A fully worked out theory of mental concepts as translucent would have to answer the following question: which aspects of a mental state such as pain do we transparently understand, and which aspects do we merely opaquely denote? It's hard to see what the answer to this question would be even

in the case of a sensory state such as pain, but it's even less clear what we could say when it comes to cognitive states such as believing that it's raining.[18]

However, even if it turns out that mental concepts are translucent, we could simply shift the focus of the argument from mental states to *those aspects of mental states that are transparently revealed to us*, call these 'mental* properties.' We can thus substitute premise 2* for premise 2**:

> **Premise 2****: I can clearly and distinctly conceive of my mental* properties existing in the absence of any neurophysiological or functional properties.

and premise 3* for premise 3**:

> **Premise 3****: If my mental* properties could exist independently of any neurophysiological or functional properties, then my mental* properties are not identical to any neurophysiological or functional properties.

It would be sufficient to refute physicalism if my mental* properties can be shown to be distinct from any functional or neurophysiological properties.

What about neurophysiological or functional concepts? It seems clear that functional concepts are transparent; a description specifying a causal role property in terms of its complete causal role completely specifies the nature of that property. The matter is slightly less clear with regards to neurophysiological properties. When a neurologist talks about 'c-fibres' is she talking about a state of the brain whose nature is entirely captured by what neuroscience has to tell us about that state, or she talking about a state *picked out* by what brain science has to tell us about it, but which may have a nature that goes beyond what brain

science can tell us about it? If the latter then premise 2* is false, as neurophysiological concepts turn out to be translucent or opaque, and hence brain science cannot afford us a clear and distinct conception of the properties of the brain.

This dispute seems to me to be largely terminological. No doubt the brain scientist, with more important matters to attend to than metaphysics, employs concepts which are indeterminate between these two options. It is up to us to decide what we mean by our talk of physical brain properties. The spirit of physicalism would seem to dictate that we define our talk of physical brain properties such that their nature can be entirely captured by neuroscience, or by neuroscience in conjunction with more basic physical sciences.[19] At any rate, if we can get an argument that refutes the kind of 'physicalist' who believes that the nature of mental properties can be fully revealed by the physical sciences then we have a significant argument. I therefore stipulate that by 'neurophysiological properties' I mean those properties which are transparently revealed to us by brain science (or brain science in conjunction with more basic sciences of matter—see footnote 19).

We have therefore demonstrated that all the concepts involved in premise 2* are transparent. The physicalist might continue to insist that the conception we reach at the end of meditation two is in some way obscure, confused or incoherent. But she is obliged to show this, and until she does, we are entitled to suppose that it is clear and distinct, as indeed it seems to be.

I take it that premise 3* is almost entirely uncontroversial, and hence we have a sound argument, from the resources of the *Meditations*, for the falsity of physicalism understood as the view that mental (or mental*) properties are either (i) identical with properties the nature of which is entirely revealed to us by neurophysiology, or (ii) identical with functional properties that are realised by properties the nature of which is entirely revealed to us by neurophysiology:

Premise 1: Anything I can clearly and distinctly conceive of is possible.

Premise 2*: I can clearly and distinctly conceive of my mental (or mental*) properties existing in the absence of any neurophysiological or functional properties.

Conclusion 1: My mental (or mental*) properties could exist in the absence of any neurophysiological or functional properties.

Premise 3*: If my mental (or mental*) properties could exist independently of any neurophysiological or functional properties, then my mental properties are not identical with any neurophysiological or functional properties.

Conclusion 2: My mental (or mental*) properties are not identical with any neurophysiological or functional properties.

3. The non-physicality of thought

Most anti-physicalist arguments of the past eighty years have tried to demonstrate that *conscious states*, defined as being states such that there is something that it is like to be in them, are distinct from physical or functional states. It was for a long time generally accepted that a functionalist account of *cognitive states* is satisfactory. There have of late been a growing number of philosophers arguing that cognitive states are in fact identical with, or grounded in, conscious states.[20] If this is the case, and if we have sound arguments for anti-physicalism about conscious states, it would seem to follow that we should be anti-physicalists about cognitive states.

However, it remains extremely difficult to settle the matter of whether cognitive states, such as thinking that it's raining, count as states such that there is 'something that it's like' to be in them. There may even be no fact of the matter as to whether this phrase from Thomas

Nagel,[21] which has its paradigm extension in the sensory realm, has application in the realm of thought.

The Cartesian argument directly supports the non-physicality of cognitive states, without relying on the thesis that they are states of consciousness. At the end of the second meditation – when I have doubted the entire physical world – I am not only conceiving of myself as a thing with sensory states; I am also conceiving of myself as a thing that thinks, that doubts, that is willing or unwilling: I am conceiving of myself as a thing with propositional attitudes. All of these states can be clearly and distinctly conceived of in the absence of anything physical or functional, and hence the Cartesian argument demonstrates that all of these states are non-physical.

The Cartesian argument shows, therefore, that we have not only a 'hard problem' of consciousness, but a hard problem of mentality in general. The mind-body problem just got tougher.

References

Armstrong, D. 1968. A Materialist Theory of Mind, London: Routledge and Kegan Paul.

Bayne, T. and Montague, M. (Eds.) 2011. *Cognitive Phenomenology*, New York: Oxford University Press.

Descartes, R. 1645. *Meditations on First Philosophy* 1996; reprinted in *Meditations on First Philosophy* J. Cottingham, (ed.) revised edition, Cambridge: Cambridge University Press.

Goff, P. MS. *Consciousness and Fundamental Reality*.

Goff, P and Papineau, D. Forthcoming. 'What's wrong with strong necessities?' *Philosophical Quarterly*.

Kriegal, U. 2013. *Phenomenal Intentionality*, New York: Oxford University Press.

Lewis, D. 1966. An Argument for the Identity Theory. Journal of Philosophy, 63:1, 17–25.

Lewis, D. 1970. How to Define Theoretical Terms. Journal of Philosophy, 67:13, 427–46.

Lewis, D. 1980. 'Mad Pain and Martian Pain,' in *Readings in the Philosophy of Psychology*, Edited by: Block, N. Vol. I, 216–22. Harvard: Harvard University Press.

Lewis, D. 1994. 'Reduction of Mind,' In Companion to the Philosophy of Mind, Edited by: Guttenplan, S. 412–31, Oxford: Blackwell.

Loar, B. 1990. Phenomenal States. Philosophical Perspectives, 4: 81–108.

Nagel, T. 1975. 'What's it like to be a bat?', *The Philosophical Review* 83.

Papineau, D. 2002. Thinking about Consciousness, Clarendon Press: Oxford.

Perry, J. 2001. Knowledge, Possibility and Consciousness, Cambridge, MA: MIT Press

Schroer, R. 2010. 'Where's the Beef? Phenomenal concepts as both demonstrative and substantial,' *Australasian Journal of Philosophy*, 88: 3, 505-22.

Tye, M. 1995. Ten Problems of Consciousness: A Representational Theory of the Phenomenal Mind, Cambridge, MA: MIT Press.

[1] To make the claim slightly more carefully, for each proposition that you can grasp you would be able to work out its truth value.

[2] Perhaps someone who had never seen red or tasted a lemon would not have a mental concept of what it's like to see red or taste a lemon. But it seems that even if you did have a full concept of, say, what it's like to taste a lemon, perhaps gained through tasting lemons whilst blindfolded, you would not be able to know from a neurophysiological description of a lemon experience that it satisfied that concept.

[3] Armstrong 1968, Lewis 1966, 1970, 1980, 1994.

[4] Descartes classes sensory experiences as a kind of thought.

[5] Lewis 1980.

[6] Lewis 1980.

[7] Descartes 1645: 18.

[8] Descartes 1645: 54.

[9] Perhaps it is slight exaggeration to say that there are *no* modal implications of the conceivability of water with a chemical composition other that H_2O. The important point is that, because of the opacity of the concept 'water', we cannot infer from the conceivability to the possibility of this state of affairs.

[10] I outline such a reduction in more detail in Goff and Papineau forthcoming and Goff MS.

[11] Descartes 1645: 110-2.

[12] Descartes 1645: 109.

[13] See Loar 1990, Papineau 2002, Perry 2001, Tye 1995.

[14] Descartes 1645: 19.

[15] Descartes 1645: 71.

[16] Descartes 1645: 72.

[17] In fact, I am inclined to think that even empirical investigation won't reveal the essence of wine, as observation reveals only the extrinsic features of material things. But the important point here is the negative one that our ordinary concept of wine does not reveal the essence of its referent.

[18] Robert Schroer (2010) offers a view according to which concepts of sensory states reveal the internal structure of those states, but opaquely denote the atomic elements involved in that structure. The view that only structural aspects of my nature are revealed to me in the conception I have of myself at the end of the second meditation, seems to me not that much more plausible than the view that no aspects of my nature are revealed to me in the conception I have of myself at the end of the second meditation. More importantly, this model cannot be applied to cognitive states.

[19] A natural view would be that neuroscience reveals brain states to be essentially constituted of certain more basic physical elements, and that the essences of those more basic elements are revealed by more basic sciences of matter.

[20] For a good range of essays on both sides of the debate, see Bayne and Montague 2011 and Kriegal 2013.

[21] Nagel 1975.